

МЕТОДИКА ОЦЕНКИ СТАБИЛЬНОСТИ ГРАММАТИЧЕСКИХ СВОЙСТВ

© 2009 г. В. Д. Соловьев, Р. Ф. Фасхутдинов

В статье сопоставляются различные методики оценки стабильности грамматических признаков. Все рассматриваемые методики восходят к ранним идеям Гринберга и реализуют их с использованием компьютерных типологических баз данных. Это позволяет получить объективную числовую оценку стабильности признаков. Анализируются причины расхождения оценок стабильности по разным базам данных.

Various methods of evaluation of grammar features stability are compared in this paper. All these methods arise from Greenberg's early ideas and realize them with the use of computer typological databases. This allows to obtain an objective qualitative estimation of features stability. Reasons for disagreement between estimations by different databases are analyzed.

Введение

Вопрос о стабильности языковых черт давно привлекал внимание исследователей. Было понято, что разные языковые черты имеют разную степень стабильности и наибольший интерес представляют стабильные черты. Почти век назад Сепир писал по этому поводу [1, p. 172]:

The general drift of language has its depths. At the surface the current is relatively fast. In certain features dialects drift apart rapidly. By that very fact these features betray themselves as less fundamental to the genius of the language than the more slowly modifiable features in which the dialects keep together long after they have grown to be mutually alien forms of speech.

(Изменение языков имеет свою глубину. На поверхности изменения относительно быстрые. В некоторых аспектах диалекты удаляются друг от друга быстро. Но в действительности эти черты являются менее фундаментальными в природе языка, чем более медленно изменяющиеся признаки, которые остаются общими для диалектов еще долго после того, как они превращаются в невзаимопонимаемые формы речи.)

Выделение стабильного ядра лексикона лежит в основе глоттохронологии Сводеша [2]. В типологической литературе встречаются утверждения о том, что те или иные конкретные типологические свойства являются стабильными, либо нестабильными. Однако эти утверждения носили, скорее, интуитивный характер и не подкреплялись значительным объемом эмпирических данных, обработанных по хорошо обоснованной методике. Для продвижения в этом направлении требовалось строгое определение стабильности и методика ее измерения.

Определение стабильности было предложено в работе [3]. В ней под стабильностью признака пони-

мается вероятность того, что в данном языке этот признак останется неизменным в течение условного промежутка времени, скажем, 1000 лет. Данное определение, разумеется, не является строгим, т.к. термин “вероятность” в нем используется, скорее, в неформальном, обыденном понимании этого слова, а не в строгом математическом смысле. Применить для оценки стабильности конкретных черт это определение напрямую件 невозможно, т.к. вероятности в математическом смысле требуют повторяемости ситуаций, что в принципе невозможно в языке, равно, как и в других аспектах человеческой культуры и истории.

Для того, чтобы иметь возможность сравнивать стабильность различных черт, необходима процедура измерения стабильности, т.е. приписывание каждой черте численного значения ее стабильности. Любая формализация меры стабильности будет одновременно и строгим определением стабильности.

Возможно, впервые подход к измерению стабильности предложил Гринберг [4, p. 76]:

If a particular phenomenon can arise very frequently and is highly stable once it occurs, it should be universal or near universal <...>. If it tends to come into existence often and in various ways, but its stability is low, it should be found fairly often but distributed relatively evenly among genetic linguistic stocks <...>. If a particular property rarely arises but is highly stable when it occurs, it should be fairly frequent on a global basis but be largely confined to a few linguistic stocks <...>. If it occurs only rarely and is unstable when it occurs, it should be highly infrequent or nonexistent and sporadic in its geographical and genetic distribution <...>.

(Если некоторый феномен может возникать очень часто и является очень стабильным, он дол-

жен быть универсальным или почти универсальным <...>. Если он имеет тенденцию появляться часто и разными способами, но его стабильность низкая, то он должен часто встречаться в языках, причем с относительно равномерным распределением среди языковых семей <...>. Если свойство возникает редко, но является высокостабильным, то такое свойство будет достаточно частотным в глобальном масштабе, но по большей части будет присутствовать в немногих семьях <...>. Если свойство редко возникает и является нестабильным, то оно должно быть вообще не присутствующим в языках или очень редким и распределенным географически и генеалогически случайным образом <...>.)

В этой формулировке представлена ключевая идея сопоставления распространенности признака среди родственных и среди неродственных языков. Большинство дальнейших исследований опирается на эту идею и развивает и уточняет ее.

Первые количественные исследования стабильности были проведены Николс [5]. Она предложила несколько вариантов мер стабильности. К сожалению, она не располагала большой типологической базой данных, что помешало широкой проверке (на большом множестве языков и признаков) и распространению ее подхода. В дальнейшем к проблеме стабильности обращались и другие авторы: Дал [6], Маслова [7], Вихман и Камхолц [8], Беляев [9]. Однако общепринятой меры стабильности грамматических признаков пока нет.

В нашей статье описаны несколько мер стабильности и проведено их сопоставление на материале базы данных “Языки мира” (далее, сокращенно, ЯМ). Одна из мер применяется к базе данных WALS [10], и стабильности сравниваются по общим базам данных. Полученные численные значения стабильности сравниваются с качественными оценками, представленными в литературе.

1. Меры стабильности

Основные сравнения проводятся на материале серии ЯМ, созданной в Институте языкознания РАН по материалам одноименной серии монографий. ЯМ содержат описания 315 языков Евразии по 3821 признаку. Признаки охватывают практически все разделы грамматики. Признаки бинарные, т.е. считается, что язык либо обладает некоторым свойством, либо нет. ЯМ доступна в Интернете по адресу www.dblang2008.narod.ru и подробно описана в [11].

Будут рассмотрены 4 меры стабильности. Учитывая бинарность подхода в ЯМ, мы ограничимся характеристикой стабильности только бинарных признаков.

При этом оценивается одновременно стабильность и признака (значение 1 в базе данных), и его

отсутствия (значения 0). Таким образом, мера стабильности характеризует вероятность одновременно и потери, и возникновения признака, т.е. его изменчивости.

Первая из мер стабильности наиболее простая в идейном и в вычислительном плане. Она восходит к одной из мер, предложенных Николс (мера 3 в [5]), и лишь несущественными техническими деталями отличается от меры В из работы [3]. Вторая является ее усложнением. Она была предложена и изучена в работе Вихмана с соавторами [3] и близка мере (2) Николс [5]. Третья мера предложена Масловой [7]. Четвертая мера является единственной, реализующей диахроническую идею подсчета числа изменений признака в ходе эволюции. Эта мера близка к идее Дала [6], но он оперирует только с установленными сравнительно-историческим методом данными, которых явно недостаточно для получения сколько-нибудь надежных результатов по стабильности. Мы реализуем подход, при котором используется филогенетический алгоритм реконструкции праязыковых состояний.

Все рассматриваемые меры объединяет то, что они оперируют с языковыми группами, родство которых надежно установлено и которые имеют примерно одинаковую временную глубину в 3–4 тыс. лет. Это роды (genets), предложенные Драйером в [12] и положенные в основу классификации языков в WALS. Они соответствуют ветвям крупных семей в более привычной терминологии.

1.1. Мера стабильности 1.

В основе этой меры лежит следующая идея. Если в большинстве языков из данной группы родственных языков (рода) присутствует некоторый признак, то наиболее вероятно, что он присутствует и в протоязыке этой группы. Столь же вероятно, что в тех языках группы, в которых он присутствует, он сохранился из протоязыка. Разумеется, есть и контрпримеры, когда признак в ходе эволюции исчезал, а потом появлялся вновь. Однако подобные обратные мутации (это перевод термина *back mutation*, который пришел из эволюционной биологии, и сейчас используется и в эволюционной лингвистике, см. работы Варноу с соавторами [13]) весьма редки. Естественно считать, что чем в большем числе языков группы признак присутствует, тем он стабильнее. Симметричным образом обстоит дело с отсутствием признака.

Исходя из этого вычислим меру стабильности признака f следующим образом. Для каждого вида V подсчитаем долю языков этого рода, обладающих признаком f . Обозначим ее $d_V(f)$. Затем вычислим $c_V(f) = |0.5 - d_V(f)|$. Эта величина показывает, насколько отклоняется от среднего значения частота признака (либо его отсутствия). Чем она больше, тем признак стабильнее (в пределах вида V). На-

конец, сумма по всем видам $c(f) = \sum_v c_v(f)$ и является мерой стабильности 1 признака.

Это определение обладает следующим недостатком – оно учитывает только один уровень иерархии родства. Рассмотрим следующий гипотетический пример. Пусть есть две семьи $S1$ и $S2$, причем обе они содержат по 2 рода: $S1 = \{R1, R2\}$, $S2 = \{R3, R4\}$. И пусть есть два признака $f1$, $f2$, причем $f1$ присутствует во всех языках родов $R1$ и $R3$ и отсутствует во всех языках родов $R2$ и $R4$, а признак $f2$ присутствует во всех языках родов $R1$ и $R2$ и отсутствует во всех языках родов $R3$ и $R4$. Мера стабильности 1 обоих этих признаков одинакова и равна 2. Однако признак $f2$ естественно считать более стабильным, так как он имеет одно и то же значение в пределах целых семей. Эти значения могут быть отнесены к протоязыкам семей $S1$ и $S2$ и их значения, вероятно, не менялись на протяжении всей истории эволюции семей. Признак же $f1$, какое бы значение он не имел в протоязыках семей, менял свое значение на этом временном интервале как минимум 1 раз в каждой семье.

Напрашивается вывод, что более адекватная мера стабильности должна учитывать несколько уровней иерархии родства. Таковой является мера стабильности, представленная в следующем пункте и учитывающая 2 уровня.

1.2. Мера стабильности 2

Для данного признака f и рода V вместо доли языков, обладающих признаком f , найдем долю пар языков рода V , таких, что признак f либо присутствует, либо отсутствует в обоих языках пары (из всего множества пар языков из V). Затем найденная величина усредняется по всем родам пропорционально числу языков в роде. Обозначим ее R (для краткости обозначений опустим указания на f и V). Далее рассмотрим множество всех пар неродственных языков (т.е. из разных семей, гипотезы дальнего родства не учитываем) и, проделав с ним аналогичные вычисления, получим число, обозначенное U . Мера стабильности 2 подсчитывается по формуле: $S = (R - U)/(1 - U)$. Определение приведено в соответствии с [3], где можно найти дополнительные пояснения. Согласно [3], идеи, заложенные в этом определении, восходят к работам [14] в области лексикостатистики, а также используются и в других исследованиях [15].

В [3] показано, что мера 2 точнее отражает стабильность признаков, чем мера 1, однако, сравнение проводилось только на искусственно порожденных данных.

Мера 2 также обладает некоторыми недостатками. Рассмотрим снова гипотетический пример. Пусть семья $S1$ содержит единственный род из 4-х языков, три из которых обладают признаком f . Семья $S2$ имеет такую же структуру. Тогда легко

подсчитать, что $R = 1/2$, $U = 10/16$. Т.е. $R < U$ и S окажется отрицательной величиной. Отрицательные значения S действительно имели место в расчетах на реальных данных [3]. Ясно, что вероятность не может быть меньше нуля, что в какой-то мере подрывает доверие к используемой формуле. В [3] отрицательные значения интерпретируются лишь как указание на нестабильность признака.

При применении меры 2 к реальным данным получаются и другие странные результаты. Рассмотрим расчеты меры 2 на основе WALS для признаков 81: “Порядок S, V, O ” и 82: “Порядок S и V ”. Оба значения признака 83 “SV” и “VS” оценены как стабильные. Одновременно с этим значения признака 81 “SOV” и “SVO” оценены как очень стабильные (т.е. более стабильные, чем “SV” и “VS”). Однако более сложная система, состоящая из, как минимум, двух элементов, не может быть стабильнее, чем ее подсистема, состоящая только из одного элемента. Любое изменение порядка SV на VS или наоборот приведет к изменению порядков SVO и SOV.

Наличие таких странных результатов для меры 2 в значительной степени подрывает доверие к ней.

1.3. Мера стабильности 3

Элегантную модель предложила Маслова [7]. Ее модель также восходит к идеям Гринберга [16]. С мерой 2 ее сближает подсчет числа пар языков, однако будут рассматриваться только пары близкородственных языков. В отличие от мер 1 и 2 в формулах Масловой напрямую используется частота рассматриваемого значения признака. Приведем необходимые определения по [7].

$P_{ij}(f)$ – вероятность перехода языка от значения i признака f к значению j за условную единицу времени t . В дальнейшем все рассуждения относятся к произвольному признаку f , упоминания о нем будут опускаться. Временной интервал t в работе [3] равен 1000 лет, что приблизительно равно времени расхождения пары языков от общего предка. Для любого i имеет место равенство $p_{i0} + p_{i1} = 1$, т.е. говоря математическим языком, имеем модель марковского процесса с 2 состояниями в дискретном времени. Предполагается, что значения вероятностей перехода зависят только от признака, но не от языка, т.е. являются своего рода универсалиями.

Из теории марковских процессов известно, что вероятность пребывания в состоянии i с течением времени стремится к пределу, называемому стационарной вероятностью и обозначаемому s_i . Имеют место равенства: $s_1 + s_0 = 1$, $s_1/s_0 = p_{01}/p_{10}$.

Интересующие нас значения – вероятности p_{01} и p_{10} – как раз и характеризуют стабильность соответственно значений 0 и 1. В принципе, чтобы найти эти величины, нужно знать значения признака в большом числе языков в предшествующие моменты времени. Поскольку это нереаль-

но, Маслова предлагает приближенный метод, позволяющий аппроксимировать искомые вероятности на основе значений признака только в современных языках.

Основная идея состоит в том, что для многих пар близкородственных языков значения признака в предшествующий момент времени (напомним, условный) должны совпадать, т.к. это был просто один и тот же язык.

Рассмотрим множество M (современных) языков, описания которых имеются. Пусть h – частота значения 1. Вероятность q значения 1 в следующий момент времени определяется формулой:

$$(1) q = hp_{11} + (1 - h)p_{01}.$$

Допустим, что для двух языков значения признака в некоторый момент времени совпадали, и подсчитаем вероятность того, что в следующий момент времени они будут различаться. Искомую вероятность назовем коэффициентом дивергенции и обозначим буквой d .

$$(2) d = 2hp_{11}p_{10} + 2(1 - h)p_{00}p_{01}.$$

После несложных преобразований получим:

$$(3) d = aq + b, \text{ где } a = 2(p_{10} - p_{11}), b = 2p_{01}p_{11}.$$

Если знать величины d и q , то по одному уравнению (3) определить величины a и b однозначно невозможно. Однако, если известно несколько пар значений (d, q) , то стандартными методами линейной алгебры можно приближенно определить величины a и b , а затем по ним и вероятности переходов p_{ij} .

Построим по множеству языков M множество M^* пар наиболее близкородственных языков следующим образом. Зафиксируем некоторую генетическую классификацию языков. У Масловой это Ethnology [17], у нас будет использована классификация, изложенная в [18] и отличающаяся от классификации [17] незначительными деталями. Выбираем произвольный язык L и находим в выбранной классификации наименьшую группу родственных языков, содержащую L . Если она содержит единственный язык, т.е. L изолят, то он исключается из рассмотрения. В противном случае выбираем в этой группе произвольную пару языков, включаем ее в M^* и исключаем эти языки из дальнейшего рассмотрения. Процедура продолжается до тех пор, пока все языки из M не будут рассмотрены. Множество языков, вошедших в пары из M^* , обозначим M' .

Величина q аппроксимируется частотой значения 1 в множестве M' , а величина d – частотой пар языков из M^* , в которых признак принимает различные значения. Для того, чтобы получить несколько пар значений величин (d, q) , требуется разбить все множество M^* на соответствующее число подмножеств приблизительно одинакового размера. M^* можно разбивать различными способами, но из линейной алгебры известно, что желательно (для

повышения точности), чтобы объекты из получаемых подмножеств, по возможности, отличались друг от друга.

В работе Масловой рассмотрено только 3 признака и для каждого строится свое разбиение множества M^* . Учитывая, что мы оперируем тысячами признаков, строить для каждого из них отдельное разбиение – весьма трудоемкая процедура. Поэтому в данной работе принято решение использовать единое разбиение следующего вида: {индоевропейские языки, алтайские и уральские языки, все остальные} и рассчитывать стабильность признаков, используя это единое разбиение множества M^* .

Как оказалось, мера Масловой также может принимать отрицательные значения, однако это происходит очень редко.

1.4. Мера стабильности 4

Стабильность признаков можно было бы подсчитать непосредственно, если бы имелись описания праязыков всех промежуточных уровней на достаточно большую глубину. При отсутствии этих данных можно попытаться их аппроксимировать с помощью математического алгоритма Maximal Parsimony, позволяющего по заданному дереву эволюции и описанию языков в его листьях реконструировать описания во всех промежуточных вершинах, вплоть до корня.

Реконструкция осуществляется исходя из принципа минимизации числа языковых изменений (мутаций), т.е. ищется такое описание всех промежуточных языковых состояний, чтобы общее число мутаций было минимально. Разумеется, ошибки возможны, реальная эволюция могла идти иным путем. Все же, в итоге, видимо, получается достаточно хорошее приближение. Этот метод хорошо зарекомендовал себя в ряде случаев, в частности, он был успешно применен для реконструкции эволюции папуасских языков [19].

После реконструкции всех промежуточных состояний нетрудно подсчитать для каждого признака общее число изменений, что и будет показателем его стабильности. В данном исследовании в качестве дерева эволюции выбрано дерево из [18].

2. Сопоставление мер стабильности

Для изучения введенных мер проведен ряд численных расчетов на материале базы данных ЯМ. Для каждого признака подсчитаны значения всех четырех мер стабильности. Для ряда признаков полученные значения не являются достоверными. В [3] для меры 2 считается, что при значениях $U > 0.9$ результаты подсчетов недостоверны. Мы также исключаем их из рассмотрения.

К недостоверным отнесены и признаки с отрицательным значением меры Масловой. Кроме то-

Таблица 1а. Сравнение мер 1 и 2

1	2	3	4
54	29	17	26
39	27	32	28
30	38	26	32
3	32	51	39

Таблица 1б. Сравнение мер 1 и 3

1	2	3	4
41	13	19	53
36	22	31	37
33	38	36	19
16	53	40	16

Таблица 1в. Сравнение мер 1 и 4

1	2	3	4
109	17	0	0
17	86	23	0
0	22	81	23
0	1	22	102

Таблица 1г. Сравнение мер 2 и 3

1	2	3	4
83	24	13	6
23	44	37	22
15	33	47	31
5	25	29	66

Таблица 1д. Сравнение мер 2 и 4

1	2	3	4
51	42	28	31
27	22	40	37
17	31	34	44
5	31	24	39

Таблица 1е. Сравнение мер 3 и 4

1	2	3	4
38	38	31	19
14	23	39	50
22	28	38	38
52	37	18	18

го, из общих соображений ясно, что для признаков с крайне низкой частотой любые подсчеты будут недостоверны. Что касается признаков с очень высокой частотой, то они являются универсалиями или статистическими универсалиями. Очевидно, что они очень стабильны. Однако их отрицания малочастотны и, соответственно, рассматриваемые методы, по меньшей мере, меры 1 и 2, не дают для них достоверных результатов. А так как значения мер 1 и 2 для признака и его отрицания совпадают, то приходится считать недостоверными и значения этих мер для высокочастотных признаков. Таким образом, часть признаков исключена из рассмотрения. Для дальнейшей обработки выделена группа из 504 признаков, для которых все значения стабильности можно считать надежными.

2.1. Корреляция мер

Прежде всего, естественно поинтересоваться, какие из этих мер и в какой степени коррелируют друг с другом. Далее, по аналогии с работой [3], для каждой из четырех мер все признаки упорядочиваются по величине и делятся на 4 примерно одинаковые по размеру группы. Это позволяет получить качественную оценку стабильности признаков. В 1-ой из групп признаки являются очень стабильными, во 2-ой – стабильными, в 3-ей – нестабильными и в 4-ой – крайне нестабильными. Такое деление не является бесспорным, и выбрано в данной работе с целью прямого сопоставления с результатами [3].

Для сопоставления мер подсчитаем число совпадающих признаков для каждой пары мер и для каждой из 4 групп. Результаты приведены в таблице 1. На пересечении строки i и столбца j указано, сколько признаков входит в пересечение групп i и j соответствующих мер.

Высокую степень корреляции демонстрируют меры 1 и 4 – соответствующая матрица является почти диагональной. Эта корреляция может быть объяснена следующим образом. Мера 4 основана на алгоритме Maximal Parsimony, который при восстановлении значения признака для языка некоторой родственной группы языков использует информацию фактически одного типа – частотность этого значения в языках данной группы. Частотность значения признака среди языков различных родов является и единственным видом информации в мере 1.

В отличие от мер 1 и 4, меры 2 и 3 используют несколько иную информацию. Мера 2 – частотность пар родственных и неродственных языков с одинаковым значением признака, мера 3 – частотность значения признака во всей изучаемой популяции и частотность пар родственных языков с разными значениями признака. Мера 3 отличается от остальных мер также специальным выбором пар родственных языков и различением значения признака и его отрицания.

Таблица 2. Сравнение 4 мер стабильности по ЯМ

Признак в литературе	Признак в ЯМ	Мера 1	Мера 2	Мера 3	Мера 4	Стабильность по ЯМ
Сингармонизм гласных [4, 20]	522: сингармонизм	7.87 (2)	0.66 (1)	0.15 (1)	31 (2)	1–2
Род [5]	1150: род	9.21 (1)	0.76 (1)	0.11 (1)	20 (1)	1
Противопоставление единственности/множественности в словоизменении именных групп [5]	1245: структура числа > единственное/множественное	8.67 (2)	0.43 (1)	0.13 (1)	25 (1)	1
Строй языка [5, 21]	3639: номинативный	8.23 (2)	0.63 (1)	0.31 (1)	39 (2)	1–2
	3648: эргативный	9.90 (1)	0.76 (1)	0.04 (1)	8 (1)	1
Наличие и место согласования с первым аргументом [5]	3713: субъектное согласование	5.64 (4)	0.26 (2)	0.31 (1)	73 (4)	?
Предлоги/послелогои [5, 20]	2885: предлоги	8.73 (1)	0.66 (1)	0.20 (1)	28 (1)	1
	2887: послелогои	8.99 (1)	0.71 (1)	0.16 (1)	22 (1)	1
Пассив [5]	1777: пассив	6.72 (3)	0.33 (2)	0.36 (1)	52 (3)	?
Тоны [5, 21]	482: тоны	9.74 (1)	0.39 (1)	0.49 (2)	22 (1)	1
Порядок SVO [20, 21]	3656: SVO	8.54 (2)	0.63 (1)	0.18 (1)	30 (2)	1–2
Порядок SOV [20]	3657: SOV	7.15 (3)	0.47 (1)	0.17 (1)	46 (3)	?
Оппозиция инклюзив/эксклюзив [5]	2551: инклюзив/эксклюзив	8.73 (1)	0.41 (1)	0.12 (1)	21 (1)	1

В итоге меры 2 и 3 оказываются в большей степени похожи друг на друга, чем на 1 и 4, хотя они близки между собой и не в такой степени, как 1 и 4. Это следует из сопоставления вышеприведенных матриц. В матрице сравнения мер 2 и 3 элементы, стоящие на главной диагонали, являются наибольшими в своих строках и столбцах, что и означает относительно более полное совпадение классов 1–4, определяемых по данным мерам. В матрицах сравнения мер 1 и 4 с мерами 2 и 3 указанное свойство не выполняется, и числа располагаются хаотично. Таким образом, меры 2 и 3 основываются на иной информации, по сравнению с мерами 1 и 4.

2.2. Сравнение с данными в публикациях

Следующим шагом в изучении имеющихся мер является сравнение полученных значений стабильности с данными, описанными в литературе. В таблице 2 собраны признаки, относящиеся в литературе к стабильным. В последнем столбце приводится итоговая оценка (в духе Николс [5]) стабильности признаков по базе данных ЯМ с учетом всех четырех мер. При близких значениях по всем мерам выбирается наиболее частое значение, при значительных расхождениях – итоговое значение не определено.

Легко заметить, что в большинстве случаев характеристика признаков, данная ранее в литературе, подтверждается. Можно обратить внимание на следующее. В большинстве случаев признаки, описанные в литературе как стабильные, получили ко-

личественные оценки даже не просто “стабилен”, а “очень стабилен”. Это можно интерпретировать так, что ранее в литературе рассматривались, в основном, бесспорные случаи.

Из введенных мер стабильности наиболее точное совпадение с ранее известными результатами дает мера 3 (Масловой) и почти такие же результаты мера 2 (Вихмана). Почти во всех случаях (кроме признака “тоны”) меры 1 и 4 классифицируют признаки как в той же степени стабильные, как и по мерам 2, 3 или же менее стабильные.

Рассмотрим более внимательно случаи явного рассогласования. Это признаки 3713: “субъектное согласование”, 1777: “пассив”, 3657: “SOV”. В первых двух из этих случаев мера 2 относит признаки ко второй группе (стабилен), а не к первой. Мера 3, хотя и относит эти признаки к первой группе, но если сравнивать абсолютные значения, то оказывается, что они для этих признаков будут наибольшими среди признаков первой группы, приближаясь ко второй группе. Другими словами, меры 2 и 3 также оценивают эти признаки, как менее стабильные, чем остальные. Отнесение их к конкретной группе является условным, определяемым выбранным разбиением на 4 группы. Возможно, что для каждой меры имеет смысл выбирать свое разбиение на группы стабильности.

Таким образом, степень согласованности между всеми рассмотренными мерами весьма высокая. Наиболее противоречивый случай “SOV”, требующий более детального изучения.

Рассмотрим еще несколько признаков, не включенных в данную таблицу. Во-первых, в литературе встречаются утверждения о стабильности некоторых черт языков в пределах определенного региона. Примером может служить утверждение, что агглютинативная суффиксальная морфология стабильна в северной Евразии [21, p. 286].

Признак 1084 “Агглютинативные языки” в БД ЯМ наиболее близок к данному свойству. Меры 1 и 4 дают для него значение – “очень стабилен”, а меры 2 и 3 – “стабилен”. Таким образом, это утверждение также подтверждается даже для более широкого класса евразийских языков.

Далее, в литературе встречаются и утверждения о нестабильности тех или иных признаков. К сожалению, большинство из них не имеет точных аналогов в ЯМ. Рассмотрим один из наиболее близких примеров. В [4, 16, 20] к нестабильным отнесены “Назальные гласные”. В ЯМ им соответствует признак № 27 “..назализованные/неназализованные (гласные)”. (Здесь и ниже число точек перед наименованием признака означает уровень в иерархии признаков в ЯМ, детальнее объяснено в [11].) По мерам 1, 3 и 4 он является стабильным, а по мере 2 даже очень стабильным. Отметим, что в WALS этому признаку в наибольшей степени соответствует значение 10.1: Vowel Nasalization: Contrastive nasal vowels, которое по мере 2 также оказывается очень стабильным [3]. Здесь результаты подсчетов с помощью и WALS, и ЯМ противоречат высказанным в литературе утверждениям.

2.3. Сравнение со стабильностью по WALS

В [9, 22] проведено сравнение меры 2 для ЯМ и WALS. Выделено 23 признака WALS (точнее, значения признаков), которые в ЯМ и в WALS совпадают или очень близки. Результаты приведены в таблице 3 (из [22, с. 160–161]).

Как видно, в большинстве случаев данные по стабильности признаков, рассчитанные по разным базам данных, совпадают или близки друг другу (находятся в соседних группах). Причины остальных случаев значительного рассогласования требуют отдельного изучения. Это может быть вызвано неполным совпадением признаков, ошибками в базах данных, погрешностями в методике и просто различной степенью стабильности признаков в Евразии и во всем мире. Обратим внимание на два наиболее интересных случая рассогласования, демонстрирующих возможные проблемы.

Признак 87.1. “Modifying adjective precedes noun (AdjN)” очень стабилен по WALS, а соответствующий ему признак ЯМ 3664 “...адъективное определение предшествует определяемому” очень нестабилен. Вероятной причиной этого является неточность в ЯМ, выражающаяся в том, что часть случаев этого порядка слов была отнесена автора-

ми статей в ЯМ к другому признаку, включающему в себя рассматриваемый, а именно к признаку: 3668: “...определение предшествует определяемому”. Та же мера 2 для БД ЯМ дает для этого признака значение “стабилен”.

Признак WALS 98.2 “Nominative – accusative (standard)” очень нестабилен, а признак ЯМ 3639 “..номинативный (строй языка)” очень стабилен. Вероятное объяснение этого расхождения состоит в следующем.

В WALS нет единого признака “строй языка”. Он расщепляется на три более частных признака – 98 “Alignment of Case Marking of Full Noun Phrases”, 99 “Alignment of Case Marking of Pronouns”, и 100 “Alignment of Verbal Person Marking”. Вполне естественно, что каждый из них по отдельности может быть менее стабилен, чем строй языка в целом. Строй языка (номинативный) может выражаться и в маркировке глагола, и в маркировке именных групп, и эти виды маркировок могут переходить друг в друга, обуславливая относительно меньшую стабильность признаков, выбранных в WALS, при высокой стабильности строя языка в целом.

Для сравнения с признаком WALS 98.2 уместнее было бы выбрать признак ЯМ 1359 “номинатив/аккузатив (падежи)”. Но и в этом случае соответствие признаков будет не полным. Признаку 1359 соответствуют два значения WALS: 98.2 и 99.2 “Alignment of Case Marking of Pronouns: Nominative – accusative (standard)”. Приведенные примеры указывают на трудности, возникающие при прямом сопоставлении двух баз данных.

Эти рассуждения показывают, что степень корреляции между стабильностью по WALS и по ЯМ на самом деле даже выше, чем это видно из приведенных в таблице 3 данных. Но даже в таком виде выявленная корреляция указывает на релевантность предложенных методов, а также на хорошую согласованность ЯМ и WALS.

Заключение

В работе впервые проведено прямое сравнение различных, рассматривавшихся в литературе мер стабильности грамматических признаков. Показано, что все четыре введенные ранее меры в высокой степени коррелируют друг с другом и с опубликованными данными. Все же между мерами есть определенные различия и требуются дальнейшие исследования для установления того, какая из них наиболее адекватна.

Также сравнивалась стабильность по двум основным типологическим базам данных – WALS и “Языки мира”. Оказалось, что степень корреляции между стабильностью по WALS и по ЯМ достаточно высокая. В качестве наиболее вероятной причины расхождения результатов предполагается несоответствие признаков в базах данных.

Таблица 3. Сравнение стабильности по WALS и ЯМ

Признак WALS	Стабильность	Соответствующий признак ЯМ	Стабильность
26.2. Inflectional Morphology: Predominantly suffixing	очень стабилен	3443 ..преимущественно суффиксальная	нестабилен
31.1. No gender system	очень стабилен	1150 ..род	очень стабилен
31.2 Sex-based gender system	очень стабилен	1200 ...пол	очень стабилен
33.1 Average of Plural prefix and 33.2 Plural suffix	очень стабилен	1257 ..аффиксация	стабилен
33.7. Plural word	нестабилен	1268 ..служебные слова	очень нестабилен
37.2. Demonstrative word used as marker of definiteness	очень нестабилен	2767 ...указательными местоимениями	нестабилен
37.3. Definite affix on noun	нестабилен	2745 ..именные аффиксы	нестабилен
38.5. Neither definite nor indefinite article	очень нестабилен	2740 ..определенность/неопределенность имени	стабилен
41.2. Distance Contrasts in Demonstratives: Two-way contrast	очень нестабилен	2615 ...ближний план/дальний план 2616 ...верхний план/нижний план	нестабилен
41.3. Three-way contrast	нестабилен	2619 ...ближний план/средний план/дальний план	нестабилен
54.1. No distributive numerals	стабилен	1326 ..распределительные (разделительные)	стабилен
65.1. Grammatical marking of perfective/imperfective distinction	стабилен	1842 ...совершенный/несовершенный	нестабилен
73.1. Inflectional optative present	очень стабилен	2359 ...желательность	нестабилен
81.1. Subject-object-verb (SOV)	очень стабилен	3657 ...SOV	стабилен
81.2. Subject-verb-object (SVO)	очень стабилен	3656 ...SVO	очень стабилен
87.1. Modifying adjective precedes noun (AdjN)	очень стабилен	3664 ...адъективное определение предшествует определяемому	очень нестабилен
107.1. There is a passive construction	нестабилен	1777 ...пассив	стабилен
112.1. Negative affix	стабилен	2809 ...отрицательные аффиксы	нестабилен
98.2. Nominative – accusative (standard)	очень нестабилен	3639 ..номинативный	очень стабилен
98.4. Ergative – absolutive	очень стабилен	3648 ..эргативный	очень стабилен
102.1. No person marking of any argument	очень нестабилен	3712 ..актант определяет форму предиката	нестабилен
102.2. Person marking of only the A argument	стабилен	3715по лицу	нестабилен
102.3. Person marking of only the P argument	очень нестабилен	3723по лицу	стабилен

Данные по стабильности грамматических признаков могут быть использованы в исследованиях по эволюции языков. Хотя статистически надежных временных оценок стабильности признаков пока не удастся дать, все же, вероятно, наиболее стабильные грамматические признаки возникли десятки тысяч лет назад, что позволяет достаточно глубоко проникнуть в историю развития языков.

Работа выполнена при финансовой поддержке ФАО РФ, проект № 2.2.1.1/6944 “Развитие Российского научно-образовательного центра по лингвистике им. И.А. Бодуэна де Куртенэ”.

СПИСОК ЛИТЕРАТУРЫ

1. Sapir E. Language: An Introduction to the Study of Speech. London: Rupert Hart-Davis. 1970 [1921].
2. Сводеш М. Лексикостатистическое датирование доисторических этнических контактов (на материале племен эскимосов и североамериканских индейцев) // Зарубежная лингвистика. Вып. I. М., 1999. С. 7–36.
3. Wichmann S. and Holman E. Assessing temporal stability for linguistic typological features. Munchen: LIMCOM Europa. Online at: <http://email.eva.mpg.de/~wichmann/WichmannHolmanIniSubmit.pdf>. 2009.

4. *Greenberg J.* Diachrony, synchrony and language universals // *Greenberg J., Ferguson Ch., and Moravcsik E.* (eds.). *Universals of Human Language*. Vol. III: *Word Structure*. Stanford: Stanford University Press. 1978. P. 47–82.
5. *Nichols J.* Diachronically stable structural features // *Andersen, Henning* (ed.), *Historical Linguistics*. 1993. Selected Papers from the 11th International Conference on Historical Linguistics. Los Angeles 16–20 August 1993. Amsterdam/Philadelphia: John Benjamins Publishing Company. 1995. P. 337–355.
6. *Dahl Ö.* The growth and maintenance of linguistic complexity. Amsterdam: John Benjamins, 2004.
7. *Маслова Е.* Динамика типологических распределений и стабильность языковых типов // *Вопросы языкознания*. № 5. 2004. С. 3–16.
8. *Wichmann S. and Kamholz D.* A stability metric for typological features. *Sprachtypologie und Universalienforschung*. 2009.
9. *Belyaev O.* Stability of language features: a comparison of the WALS and JM typological databases. *Proceedings of the Int. Conf. "Cognitive Modeling in Linguistics"*. FCCL. 2009. Online at: http://fccl.ksu.ru/conf_CML_2008/jm-wals-stab-2.doc
10. *Haspelmath M., Dryer M., Gil D., Comrie B.* *The World Atlas of Language Structures*. Oxford: Oxford University Press. 2005.
11. *Поляков В.Н., Соловьев В.Д.* Компьютерные модели и методы в типологии и компаративистике. Казань: КГУ, 2006.
12. *Dryer M.* Large linguistic areas and language sampling. *Studies in Language*. 13. 1989. P. 257–292.
13. *Nakhlen L., Ringe D., Warnow T.* Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Languages*. 2006. P. 382–420.
14. *Thomas D.* Basic vocabulary in some Mon-Khmer languages. *Anthropological Linguistics* V. 2. No. 3. 1960. P. 7–11.
15. *Hawkins J.* *Word Order Universals. Quantitative Analysis of Linguistic Structure*. New York: Academic, 1983.
16. *Greenberg J.* The diachronic typological approach to language // *M. Shibatani, T. Bynon* (eds.) *Approaches to Language Typology*. 1995. P. 143–166.
17. *Gordon R.* (ed.). *Ethnologue*. 15th Edition. SIL International. Online at: www.ethnologue.com. 2005
18. *Бурлак С.А., Старостин С.А.* Введение в лингвистическую компаративистику М.: УРСС, 2001.
19. *Dunn M., Terrill A., Reesink G., Foley R.A., Levinson S.C.* Structural phylogenetics and the reconstruction of ancient language history. *Science*. V. 309. No. 5743. 2005. P. 2072–2075.
20. *Croft W.* *Typology and Universals*. Cambridge: Cambridge University Press. 1996.
21. *Nichols J.* Diversity and stability in languages // *Joseph, Brian D. and Richard D. Janda* (eds.), *The Handbook of Historical Linguistics*. Oxford: Oxford University Press. Malden/Oxford/Melbourne/Berlin: Blackwell Publishing. 2003. P. 283–310.
22. *Polyakov V., Solovyev V., Wichmann S., Belyaev O.* Using WALS and *Jazyki mira*. *Language Typology*. V. 13. 2009. P. 135–165.